HOW THE AIR FORCE SHOULD STAY ENGAGED IN COMPUTER VISION TECHNOLOGY DEVELOPMENT

Mark B. Skouson, Major, USAF
April 2007

Blue Horizons Paper
Center for Strategy and Technology
Air War College

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **APR 2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE **How the Air Force Should Stay Engaged in Computer Vision Technology Development**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air University,Air War College,Center for Strategy and Technology,Maxwell AFB,AL,36112**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**43**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Acknowledgements

I am indebted to Dr. Paul Moscarelli and Dr. John Ackerman for the excellent background reading material they provided to get me started on this research. I give additional thanks to Dr. Moscarelli who, as my research advisor, provided me with much needed feedback and direction. I appreciate the time the survey respondents took to help me. In order to keep the survey anonymous, I am not able to list their names. I also wish to thank all my classmates in Flight 11 for their patience with my questions about formatting, wording, and endnote help. Special thanks go to Rae Agren, Ellie Robinson, and Jeff McLemore for finding some of my mistakes. Of course, any remaining mistakes are my fault not theirs. Lastly, I wish to thank my wife and children for their patience and understanding throughout my ACSC experience.

Abstract

The Air Force must continue to play an active role in shaping future computer vision technologies by investing in sensors networks, data fusion, technology transition, and artificial intelligence. A survey shows that while that computer vision technology appears to be progressing in general agreement with Air Force needs for the 2030 timeframe, a few gaps exist that the Air Force must address. The survey combines the judgment of 13 experts from academia and industry, and the results are compared to the Air Force's expected computer vision needs, as documented in the Air Force 2025 Study.

The survey results and accompanying analysis are a significant contribution to the military decision-making community. The results show expected maturity information for specific computer vision technologies, estimate the relative difficulty in maturing the technologies, and provide a list of technical and non-technical hurdles. The analysis also shows how specific technologies relate to possible future threats. The information is invaluable for anyone making strategic technology-related decisions.

TABLE OF CONTENTS

LIST OF TABLES

Section 1: Introduction

In the 1995-1996 academic year, Air University prepared a set of research papers in response to a directive from the Chief of Staff of the Air Force to "examine the concepts, capabilities, and technologies the United States will require to remain the dominant air and space force in the future."[1] The study was called Air Force 2025. The study found computer vision to be an important technology area. This report provides estimates for when the computer vision requirements the 2025 study calls for by surveying experts in the academic and commercial research community. An analysis of the survey data shows that the Air Force should invest in sensors networks, data fusion, improvements in technology transition, and artificial intelligence.

A major contribution of the survey is a repository of information that senior Air Force leadership can use to make decisions. This information, when coupled with a matrix of possible future threats, gives senior leadership a powerful decision making tool. Leaders can make probabilistic statements about the likelihood of specific future threats, and then use the information provided in this report to determine which technologies to invest in to combat the threat.

The remainder of this section defines computer vision for this report's purposes, explains why computer vision is important to the Air Force, and explains why the Air Force should be interested in the opinions of the academic and commercial research communities. Section two describes the survey used to gather expert responses, including how the experts were selected and the questions they were asked. Section three shows the significant advances the experts projected through 2030 and their relative difficulty. Section four compares the survey results to previously conducted study of related technologies. In addition, Section four analyzes the utility of computer vision capabilities in the context of several world threat scenarios.

Computer Vision

In their book *Computer Vision*, Shapiro and Stockman define computer vision as the study of how to "make useful decisions about real physical objects and scenes based on sensed images."[2] They explain that many of the fundamental issues inherent in computer vision can be categorized into four categories: sensing, encoded information, object representation, and algorithms.[3] Therefore, the computer vision umbrella covers a wide variety of topics from how best to capture data, to ways of extracting information (including perhaps wisdom) from that data. For some people, the computer vision umbrella also covers advances in both sensor hardware and human-computer interfaces. In the end, the goal of computer vision is in-line with the goal of most computer systems: to do for humans what they do not want to do themselves. In many cases, the desire is that the computer be faster and more reliable than humans are.

Researchers have several different methods of partitioning the computer vision research space described above. Additionally, many people consider computer vision to be a subfield of artificial intelligence. Just a few of the other names for research herein described as computer vision include pattern recognition, machine vision, image understanding, robot vision, image processing, and image analysis. Differentiating the nuances between these names is beyond the scope of this report. This research investigates all major aspects of these research areas from sensing the physical scene to articulating decisions based on the information in the scene.

---

[1] "Air Force 2025," (Air University, 1996).
[2] George Stockman and Linda G. Shapiro, *Computer Vision* (Prentice Hall PTR, 2001), 13.
[3] Ibid.

Computer Vision Technologies Required by the AF 2025 Study

Computer vision is an important technology area for the Air Force. The Air Force 2025 study ranked the two computer vision related fields—image processing and artificial intelligence—in the top 11 of 43 key technologies identified in the study.[4] Additionally, image processing and artificial intelligence technologies were critical to three and four of the 11 conceptual systems considered most important to the Air Force's future, respectively.[5]

The Air Force 2025 study expressed a need for computer vision in three general areas: collecting more image data, presenting it in usable form to aid humans in making decisions, and automatically making decisions from it. The 2025 study projects that computer vision systems will mature to the point that by 2025 they will be able to relieve humans of much of the burden of interpreting image and video data.

Automatic target recognition was one of the specific computer vision technologies in the Air Force 2025 study. As an example, consider the Worldwide Information Control System concept.[6] In this conceptual system, computer vision is used "to automatically interpret and analyze images (e.g., automatically detecting and identifying potential targets)."[7] Automatic target recognition would greatly improve the speed with which the Air Force could find, fix, and eliminate targets.

In addition to automatically recognizing targets, computer vision has application to efficient database management. In the Information Operations Architecture concept for 2025, one portion of the architecture is a "knowledge system."[8] Among other things, the "knowledge system" controls data storage, analysis, and retrieval. It "automatically recognizes gaps, deficiencies, or outdated information in the databases and, without human intervention, searches the global information net. … The architecture also reviews numerous satellite images and alerts human analysts to any changes found at potential target areas making obvious exceptions for weather."[9] This would give the Air Force the most up-to-date information to improve decision-making.

The 2025 study also projected the use of an autonomous vehicle concept called StrikeStar. The StrikeStar concept is an autonomous unmanned aerial vehicle that would contain "an artificial intelligence engine…to perform a wide range of pilot functions."[10] These functions could include takeoffs, landings, and collision avoidance. The StrikeStar would give the Air Force the option to go without pilots into very dangerous environments. The functions necessary for autonomous vehicle operations require computer vision capabilities.

The study also discussed the availability of intelligent surveillance, advances in sensor capabilities, and improved human-computer interfaces. In addition to the specific computer vision technologies mentioned, the study called for high levels of artificial intelligence within the computer vision systems. This improved computer intelligence would require improvements in computer-based visual understanding, self-configuring systems, and perhaps the ability for artistic abstraction.

---

[4] "Air Force 2025," vol. 4, ch. 3, p. 54.
[5] Ibid.
[6] Ibid., vol. 1, ch. 2.
[7] Ibid., vol. 1, ch. 2, p. 13.
[8] Ibid., vol. 1, ch. 1, p. 17.
[9] Ibid., vol. 1, ch. 1, p. 18-19.
[10] Ibid., vol. 3, ch. 13, p. 39.

Outsourcing Technology Development

The above examples have shown the conceptual applications of computer vision to the Air Force. Computer vision has multiple applications in commercial industry including medicine, quality assurance, and access control. In many instances, the technologies used in industry applications are also useful for military applications. Of course, defense and commercial industries leverage technology development from each other in building their own products. This is especially the case in areas that are highly information technology-centric.

From 1991 through 1997, the defense department began a series of initiatives to increase reliance on commercial practices and products within the technology development process.[11] The primary reasons were to accelerate product development and reduce costs. Some argue that the initiatives are resulting in a loss of expertise within the military. Others argue that much of the money saved in research and development is lost in modifications and integration. However, the Department of Defense (DoD) has shown no signs that it is planning to change course.[12]

The AF 2025 study findings and the continued DoD emphasis on commercial technology development are the primary motivations for this study. This research investigates advancements in the computer vision field because of its importance as identified by the AF 2025 study. The research investigates these advancements through the eyes of the commercial and academic research sector because of the Air Force's continued reliance on them in the development process. Since the future is uncertain, opinions and judgment play a vital role in making projections. This research makes the future projections by relying on the opinions and judgment of experts as to what technological advances will occur in the computer vision field through 2030. The next section explains how the expert judgment is gathered.

Section 2: Data Gathering Method

Studies projecting possible technological advances are not new. Several well-documented techniques exist.[13] For this research, the goals for the technique were a process for combining a group response and the ability to conduct the study without face-to-face meetings. Based on these requirements, the Delphi method was selected.

Delphi Method

The Delphi method is a surveying method developed by RAND Corporation in the 1950s as part of their continuing efforts to improve decision-making.[14] RAND designed the method to aid in problem solving in the absence of complete information. In these situations, decisions depend on opinion, wisdom, or judgment, and it is desirable to have multiple experts collaborate on the decision making process. The rational was "primarily the age-old adage 'Two heads are better than one.'"[15] The Delphi method provides a systematic process to gather and use the information gathered from these groups of experts. It attempts to improve the group response by

---

[11] Gregory Saunders, "COTS in Military Systems: A Ten Year Perspective," in *Military & Aerospace Electronics Show* (Baltimore, MD: 2004), 9.

[12] Chris A. Ciufo, "COTS: 10 Years after - Well, Sure…but What About the Next 10 Years?," *Military Embedded Systems* (2006).

[13] Jerome C. and Theodore J. Gordon Glenn, ed., *Futures Research Methodology Version 2.0* (American Council for the United Nations University Millennium Project, 2003).

[14] Chitu Okoli and Suzanne D. Pawlowski, "The Delphi Method as a Research Tool: An Example, Design Considerations and Applications," *Information and Management* 42 (2004): 16.

[15] Norman C. Dakley, *The Delphi Method : An Experimental Study of Group Opinion* (Santa Monica, Calif.: Rand Corporation, 1969), v.

using anonymity and iterative, controlled feedback.[16] Therefore, "the Delphi technique is a method of eliciting and refining group judgments."[17]

In the method, a moderator leads a group of experts through a series of anonymous surveys, controlling the feedback between rounds. The controlled feedback allows all voices to be heard, the iterative process encourages consensus, and the anonymity ensures answers are evaluated on their own merit rather than the reputation of the respondent. Researchers have used the Delphi method for such diverse tasks as determining key issues for knowledge management, identifying problems in software development, and projecting product demand. [18,19,20]

Studies show that the Delphi method generally works well. Rowe and Wright analyzed 27 Delphi studies and concluded that, on average, they outperformed statistical groups and standard interacting groups. [21] They did indicate that some advanced structured group procedures were comparable to the Delphi method.[22] The Delphi method, however, does not require face-to-face meetings. That, coupled with the solid performance, led to the selection of the Delphi method.

The Experts

The Literature indicates that the quality of the experts in a Delphi study is an important factor to obtaining good results.[23] Since the Delphi method topic is usually highly speculative, the general population "might not be knowledgeable enough to answer the questions accurately." [24] Additionally, having more experts is not always better. Based on the literature, the method works best with 10-18 experts. [25] This research followed the method for selecting experts presented by Okoli and Pawlowski, which starts by identifying the characteristics of experts needed, and uses a referral system to populate the group.[26]

It was decided that for this survey an expert should have at least 10 years of computer vision research experience in academia or industry. Out of about 35 contacts, 15 agreed to participate. Thirteen actually returned the first survey, and 11 completed the second round. Of the 13, all had a Ph.D. in computer science, electrical engineering, or related field. The experts averaged about 26 years of experience after receiving their doctoral degrees, and only one had received his Ph.D. within the last 10 years. All but two were fellows in at least one technical society. Several were fellows in multiple societies. Four were currently working in industry research positions, the rest were university faculty. Two universities were represented twice. Most had been editors of research publications. The two that dropped out after the first round were from universities and were fellows.

---

[16] Ibid.
[17] Ibid.
[18] Insu Park et al., "Guest Editorial:Part 2: Emerging Issues for Secure Knowledge Management—Results of a Delphi Study," *Systems, Man and Cybernetics, Part A, IEEE Transactions on* 36, no. 3 (2006).
[19] Sasa Dekleva, "Delphi Study of Software Maintenance Problems" (paper presented at the Conference on Software Maintenance, Los Alamitos, CA, USA, 1992).
[20] Marvin A. Jolson and Gerald L. Rossow, "The Delphi Process in Marketing Decision Making," *Journal of Marketing Research* VIII (1971).
[21] Gene Rowe and George Wright, "The Delphi Technique as a Forecasting Tool: Issues and Analysis," *International Journal of Forecasting* 15 (1999): 372.
[22] Ibid.
[23] Ibid.: 368.
[24] Okoli and Pawlowski, "The Delphi Method as a Research Tool," 19.
[25] Ibid.
[26] Ibid.

The Questions

The intent of the survey was to determine likely future capabilities of the computer vision field. In order not to bias the group, the first round of questions was open-ended. A by-product of having such an experienced set of experts was that each of them was extremely busy, so the goal was for each round of the survey to take less than 30 minutes to complete. The main question was as follows: List (at most 5) significant computer vision or image pattern analysis advancements that will occur by the year 2030. In addition to the main question, the moderator asked two more questions to aid in analysis and direct further research in this area. The second question was "What are the (at most 5) main technical hurdles that need to be overcome to reach this end state in 2030?" The last question was "What non-technical factors (i.e. cultural, economic, environmental) might have an adverse effect on the future of pattern recognition by 2030?" For each of the three questions, the moderator encouraged participants to explain their responses.

The moderator aggregated all the responses for question one and merged duplicate responses to come up with 35 unique advances that the participants expected to occur by 2030. For the second and third rounds, the moderator asked each of the participants which of the 35 advances would be mature in the near-term, mid-term, long-term, or very long-term. Near-term was defined to mean between now and 2014, mid-term was between 2015 and 2022, and long-term was between 2022 and 2030. Very long-term was for those advances that were expected to occur after 2030. A mature technology was defined as one that achieved widespread use or capability on par with human performance.

The moderator limited the scope of rounds two and three to question one, since question one was the critical question for the forecast. In a few instances, participants did not respond about all 35 advances. In other instances, the participants split their vote between two timeframes for a particular technology. For example, participants answered some questions with "near to mid-term" rather than just "mid-term." In these cases, the moderator divided the vote to put a half-vote in each category.

None of the participants revised their opinions in the third round. The intent of the Delphi method is that additional rounds lead to higher consensus within the group. However, studies have shown that for forecast studies, increased consensus with additional rounds may be difficult to achieve.[27] Another study suggested that attrition might give a false sense of consensus.[28] As will be shown in the next section, the degree of consensus achieved was sufficient to show the relative difficulty in maturing the different technologies.

Section 3: Data Analysis

As mentioned in the previous section, the survey participants identified 35 unique significant technology advances. This section discusses those advances in the context of the needs established in the Air Force 2025 Study. Although the moderator did not lead study participants to predict technologies that were discussed in the 2025 study, each of the major computer vision technologies from the 2025 study surfaced through the course of the Delphi survey.

This section discusses only those technologies from the study that directly relate to the 2025 study. Appendix A contains the author's analysis of the data that was less directly relevant to the 2025 Study. This report does not provide an in-depth description of the technology, nor does it explain the current state of the technologies. Rather it is restricted to analysis of which

[27] Rowe and Wright, "The Delphi Technique as a Forecasting Tool: Issues and Analysis," 370.
[28] Ibid.: 364.

technologies might be available and when. Appendices B through E contain the raw information collected from the survey and the tabularized results of the voting procedure. In order to maintain anonymity, the survey participants are not cited when they are quoted.

Overall, the survey panel was quite optimistic in its outlook. For 34 of the 35 technologies, the majority of the respondents thought the technology would occur by 2030. The survey also shows relative difficulty of each technology. For 17 of the 35 technology advances, all survey participants projected maturity by 2030. For this report, these are defined as the easiest technologies. For 7 technology advances, all but one participant projected maturity by 2030. These are defined as the moderately difficult technologies. For the remaining 11, at least two participants projected maturity after 2030. These will be referred to as the most difficult technologies in this report.

Of the seven technology areas highlighted from the 2025 study, only the sensor improvement area was considered an easiest technology. The moderately difficult technologies were efficient database management, autonomous vehicle operations, and human-computer interfaces. The most difficult technologies were automated target recognition, intelligent surveillance and monitoring, and high-intelligence systems. The high-intelligence systems area was clearly the hardest of the technologies. The next several paragraphs analyze the survey data in these technology areas.

Automatic Target Recognition

The survey showed automatic target recognition to be one of the most difficult technologies. As described previously, the Air Force 2025 Study identified automatic target recognition as an important part of future conceptual systems. One participant predicted that by 2030 computers would achieve human-like performance “for category-level object recognition in natural, cluttered scenes for visible (non-occluded) objects, a limited number of categories and simple spatial configurations.” Even with the restrictions of limited categories, simple spatial configurations, and no occlusion, the panelists considered this one of the most difficult technologies. When the moderator questioned the panelists on when this would be most likely to occur, two panelists said it would not mature by 2030.

One of the more optimistic participants proposed that “Specific object recognition will be very robust. It will be possible to "train" a system by showing one or more examples of the objects in an un-segmented scene and the object will be recognizable in new images taken under a broad range of conditions (pose, lighting, differing shape configurations, etc.).” The participant gave as reason for the optimism that “By 2030, it will be better understood how to describe or identify a previously unseen object in terms of a great deal of prior knowledge about a very large number of broad object classes. A large-scale ontology of objects and scenes will have been developed, and given one or more images, recognition methods will describe the image content in terms of this ontology.”

Here the survey gives the first indication that sensor networks can improve computer vision performance. The survey showed that automatic target recognition might mature by 2022 if it is able to use information from multiple sensors. Otherwise, it might not mature until 2030. As will be seen, a similar result occurs with the questions relating to automated surveillance and monitoring.

Efficient Database Management

According to the survey, efficient database management was a moderately difficult technology. Image database management, including both efficient storage and efficient retrieval, is important to the Air Force because of the volume of image data it collects and because of the

time-critical nature of the information stored. From a commercial standpoint, users want the ability to perform quick searches of huge databases. With the advent of the World Wide Web this has become less a luxury and more a necessity. However, so far text-based searches have proven friendlier than image-based searches.

The moderator asked when a system would be developed to locate and manage networked image and video content. The fully developed system should be able to locate videos of events and general locations with specific participants. The most common answer from the panelists was in the near-term with five votes. However, a few of the panelists thought this was more of a mid or long-term capability. One panelist thought it wouldn't be fully developed until after 2030.

As a related concept, an image database that could automatically annotate itself based on the semantic information from the image would be quite useful. Currently many of the commercial image retrieval systems rely on image labels or the content of surrounding text. Having a database that automatically labels the images based on surrounding text would improve retrieval speed. According to the survey, this was one of the easiest technologies to mature.

While automatic labeling of images may be a relatively easy technology in a commercial context, it would need some modification before it could be applied to many Air Force applications. Many of the Air Force images are collected from surveillance and reconnaissance assets where there is no surrounding text. In this case, the automatic labeling information would have to come from another source, so the technology more difficult to mature to a military application. The conclusion the Air Force may have to make considerable additional investment to adapt some commercial computer vision applications to military use. Intelligent transportation systems is another area where this is the case.

Intelligent Transportation Systems

The panelists considered intelligent transportation a moderately difficult technology. Autonomous driving has received considerable attention for quite some time. In 1977, Japanese engineers created a robot capable of traveling 20 mph along streets.[29] In 1995, researchers at Bundeswehr Universität München and Carnegie Mellon University independently created vehicles that completed road trips of 1000 and 3000 miles, respectively. For the Air Force, this technology has application to autonomous operation of aerial combat and transportation systems.

The panelists generally considered the technology for automated driving along long stretches highway to be easier than to automate driving in mixed traffic on public roads. A few of the panelists thought slow social acceptance would delay maturity in this technology area. One brought up the possibility that liability issues would prohibit companies from adding this technology to their products. Another panelist thought the public would be slow to embrace the technology.

As with image database management, the Air Force would have to modify this technology to adapt it to their use. For example, the steering cues and avoidance systems would likely differ for ground-based and airborne applications.

Intelligent Surveillance and Monitoring

Another technology that was prevalent in the AF 2025 study was intelligent surveillance. Since two panelists thought this general technology area would mature in the very long term, the technology should be considered a difficult one to mature. As a specific example of an intelligent surveillance application, one panelist projected that by 2030 cameras would be available to persistently watch public locations and identify present individuals. The survey

[29] Paddy Comyn, "Sensing Forward to a Driverless Future," *The Irish Times* 21 February 2007.

showed this capability as moderate difficulty, since only one panelist thought it would mature after 2030. Another participant predicted that by 2030 cameras with integrated sensing and processing units would be available and would produce not only images, but also interpretations. This would in turn “enable a number of fundamental changes, including ubiquitous robots with sophisticated vision interacting with people in everyday scenarios.” The survey put this prediction in the relatively difficult category. As with automated target recognition, the participants believed a network of sensors with associated processing would improve surveillance.

Sensor Improvements

In addition to surveillance and monitoring systems, the participants made predictions about advances in sensors. Higher quality images and video are beneficial to the Air Force because they lead higher signal-to-noise ratio which in turn improves performance of computer vision systems. One of the difficulties in transitioning this technology to military application would be converting it for use in a harsher environment than is typical for commercial devices.

All participants thought extremely large format (gigapixel) video sensors would be widely available by 2030. One participant predicted that by 2030 cameras would have automatic adaptive calibration, including photometric considerations. Again, this turned out to be one of the easiest technologies. Another specific prediction was the advent of miniature imaging sensors with on-board computation. The panelists thought this was a relatively easy technology. Half the panelists considered it a near-term technology, while the other half considered it a mid-term technology.

Human-Computer Interfaces

Several of the survey panelists projected significant advances in the area of human computer interfaces. Conceptual systems within the Air Force 2025 study highlighted the military importance of improved interfaces to enhance both the speed and accuracy of human decisions. To the general prediction of a “natural human-computer interface using vision and speech,” only one thought it would mature later than 2030, so it belongs in the moderately difficult category. One panelist predicted that by 2030 we would have stereoscopic television and internet movies. Another participant predicted that 3-dimensional displays that would not require glasses would become widely available. Both of these were also moderately difficult technologies.

Participants were more optimistic about specific limited predictions in this area. One survey participant predicted “interactive zoom/pan/tilt over television/internet for unlimited number of users of both video and audio.” The participant responses put this in the easiest technology category. Participants were also very optimistic about the widespread use of low-cost capture of human motion. This technology would allow for device free controllers such as those now offered in some video game consoles.

As computers become more prevalent in everyday items, human-computer interfaces will be more important. When asked when almost all everyday objects would have computers embedded in them, eight participants thought this was a near-term event. One thought it was mid-term, and two thought it was long-term.

High-Intelligence Systems

In addition to the specific categories just discussed, several conceptual systems in the 2025 study made a general assumption that computer vision systems would have a high level of human-like reasoning skills. Of the seven computer vision areas from the 2025 study, this one was clearly the most difficult. One participant predicted systems would be able to self-

configure to adapt to changes in its environment. The panelists categorized this as a relatively difficult technology. Another participant predicted that by 2030 we would have humanoid robots with visual understanding. Of all 35 technologies in the survey, this was the second least likely technology to mature by 2030 of all in the survey.

The least likely technological advancement of all in the survey was for visual thinking systems capable of abstraction, association and visual creativity. Eight of the 11 participants thought this would mature after 2030, with one saying it might be impossible. Of the remaining three, one stated that it was a near-term technology, citing that in some cases computer vision systems have created art. Another said it was a long-term technology, and one said it was a mid to long-term technology.

Four recommendations surface from the above analysis. The first is that the Air Force should apply additional resources to help mature technologies in the high-intelligence systems area. It was the only area the majority of participants predicted would mature after 2030. The Air Force should take caution, however, in how it applies resources to this area. In the article "Out of Their Minds," Geoffrey James notes that venture capitalists are avoiding artificial intelligence. They have learned that the investments do not usually pan out.[30]

An alternative approach to traditional funding is the use of grand challenges. Grand challenges are challenges put out to the community at large with a reward for the first to complete the challenge. The Defense Advanced Research Projects Agency (DARPA) has successfully used grand challenges to promote research in a specific area.[31] The National Institute of Standards and Technology (NIST) has also used them with success.[32] The Air Force should consider using grand challenges to promote advances in artificial intelligence, especially since traditional funding methods have had little success.

Other than the high-intelligent systems, the technologies the Air Force needs should be mature by 2030. The analysis shows that the Air Force will need to make additional investments to apply the technologies to Air Force applications. The second recommendation is that the Air Force should continue to invest in professionals that are able to facilitate technology transfer from academia and industry to the military.

A team conducted a study to determine what the officer of the future should like.[33] The team did find that the officer of the future should know how, when, and why to apply technology.[34] The Air Force should continue to press for professionals that not only know how, when, and were to apply the technology, but are able to facilitate technology transfer.

The analysis shows that networking multiple sensors will improve computer vision capability. The third recommendation is that the Air Force should apply resources to promote advances in sensor networks. The fourth recommendation follows from the third. Since the Air Force should focus on sensor networks, the Air Force should also focus on ways to combine the data from the various sensors. Data fusion is the technology to merge information together from many sources. Therefore, the fourth recommendation is to focus on data fusion.

The 2025 study found that data fusion was one of most important technologies when considering all future Air Force future system concepts. When considering the 11 most

---

[30] Geoffrey James, "Out of Their Minds," *Red Herring*, 22 August 2002.
[31] DARPA, "DARPA Grand Challenge 2005," http://www.grandchallenge.org/.
[32] NIST, "NIST Face Recognition Grand Challenge," http://face.nist.gov/frgc.
[33] Anna Simons et al., "The Military Officer in 2030: Secretary of Defense 2003 Summer Study," (Director of Net Assessment, Office of the Secretary of Defense, 2003).
[34] Ibid., 38.

important system concepts, data fusion was the most important technology of all.[35] The analysis in this report confirms the importance of data fusion.

Section 4: Additional Analysis

The preceding section analyzed the availability of computer vision technologies based on a panel of experts from academia and industry. To validate the analysis, the results were compared to a statistical modeling method.

Comparison with a Statistical Model Future Study

When sufficient historical data exists, statistical modeling (using historical data and mathematical equations to project future events) can be a reliable approach for conducting futures studies. Ray Kurzweil's recent book, *The Singularity is Near* presented a futures study that included computer vision related technologies using statistical modeling. He bases his projections on the history of growth in several technology fields. The evidence most closely related to computer vision includes exponential growth in computer technology, speech recognition, and modeling of the human brain.[36] Using this historical data, Kurzweil predicted computers would have human-like performance by 2029, and much better than humans by the 2040s.[37] Through association and analogy, he predicts the human-like performance will include human-like computer vision. The analogy is logical because computer vision performance is heavily dependent on computer processing, and the image processing algorithms are similar to speech processing algorithms.

Sadly, there is a lack of historical evidence using actual computer vision systems to predict its future statistically. This is understandable because of the lack of standardized testing procedures for computer vision applications. However, the National Institute for Standards and Technology (NIST) has taken steps that may remedy this problem. They sponsored a Face Recognition Grand Challenge that ended in 2006. The goal was a magnitude improvement (10 times better) performance over previously measured results from similar tests conducted in 2002. Preliminary results indicate that NIST has made significant progress toward (and may have achieved) this goal.[38]

Should the face recognition community be able to achieve the same 10 times improvement every four years, then by 2030, computers would be able to recognize faces with 99.998% accuracy from a group of four million possibilities (in a laboratory environment). At this point however, there is not enough history evidence to say with any confidence whether this trend could continue.

Another computer vision area that does have a longer trend history is the digital image sensor field. Mackey reported that over the last 10 years both the digital image sensor resolution (megapixels/sensor) and density (megapixels/sensor area) have been increasing exponentially, while cost per megapixel has been decreasing exponentially. He also identifies future technologies that show promise in furthering this trend. He notes, however, that the trend is fueled by consumer demand for smaller, cheaper, higher-resolution digital cameras. He sees a limit to consumer demand; at that point, the trend will level off.[39]

---

[35] "Air Force 2025," vol. 4, ch. 3, p. 54.

[36] Ray Kurzweil, *The Singularity Is Near : When Humans Transcend Biology* (New York: Viking, 2005), 56-103, 292.

[37] Ibid., 263, 96.

[38] P. Jonathon Phillips et al., "Recognition Grand Challenge Results" (paper presented at the 7th International Conference on Automatic Face and Gesture Recognition, 2006), 1.

[39] Morgan Mackey, "Nanotechnology Applications for ISR: The Solution to the Intelligence Gap? (DRAFT)" (Air University, 2007).

While Kurzweil is optimistic about computer vision's capabilities by 2030, not everyone agrees with his predictions.[40,41] From a computer vision standpoint, Kurzweil may be overstating historical successes. He claims that in 1999 machines could recognize faces.[42] NIST measured face recognition systems in 2002 at about 80% effective (in lab tests). In 2003, the Tampa Police Department found their face recognition system to be ineffective at only 61.4%. Clearly, the face recognition problem is still not solved. Kurzweil also states, "robots with no human intervention have already driven nearly across the United States on ordinary roads with other normal traffic."[43] Technically the statement is correct, but humans controlled both the brakes and the throttle,[44] and the robot was never autonomous for more than 70 miles at a time.[45]

Time will tell whether Kurzweil is overly optimistic or not; both our survey participants and Kurzweil agree that the field will experience significant advances by 2030. However, most of the participants stopped short of saying all areas of computer vision would be mature by 2030. In fact, the survey participants described some of the hurdles, both technical and non-technical, that computer vision faces.

Major Technical Hurdles

Interestingly, when the survey moderator asked the participants to list significant technical hurdles to progress there was little overlap in the answers. However, there were a few areas where the participants' responses agreed with one another. Two participants stated the need for faster computers. A related comment stressed the need to harness the power of distributed computing. A few participants mentioned the need for better models to represent objects. Two noted the need to rely on multiple types of sensors, such as, visible light sensors and infrared sensors. Two participants stated the need to improve the ability to capture structure from motion.

An interesting area of further study would be to use the Delphi method to attempt to find consensus on the most important of these challenges. The challenges may provide a leading indicator for computer vision's progress over time, since the ability to overcome these hurdles in a timely manner would serve as an indicator as to the likelihood of reaching the predicted performance. Appendix B contains the complete list of submitted responses to question two.

Other Obstacles to Progress

When asked to comment on non-technological obstacles that might hinder progress in computer vision, there was significant redundancy among the responses. The obstacle cited most was funding. Some participants indicated that social or economic factors might be the cause of reduced funding. These factors included global warming or other environmental problems and war. Another major obstacle was social acceptance. Some participants mentioned the hesitancy for humans to trust computers, while others cited privacy or racial discrimination concerns.

Finally, some participants claimed that a breakdown of Moore's law would impede progress. Moore's law is the prediction that the number of transistors on a chip will double every 18 months or so. Loosely speaking, Moore's Law infers an exponential increase in

---

[40] James, "Out of Their Minds."
[41] Harold A. Linstone et al., "Book Review and Discussion: The Singularity Is Near: When Humans Transcent Biology," *Technological Forecasting and Social Change* 73 (2006).
[42] Kurzweil, *The Singularity Is Near*, 290-91.
[43] Ibid., 286, 92.
[44] "No Hands across America Home Page," http://www.cs.cmu.edu/afs/cs/user/tjochem/www/nhaa/nhaa_home_page.html.
[45] "NHAA Journal," http://www.cs.cmu.edu/afs/cs/user/tjochem/www/nhaa/Journal.html.

computer speed and storage capacity. While Moore’s Law is technical, it is not something computer vision researchers consider within their control.

World Futures Scenarios

The participants’ list of non-technical hurdles highlights how non-technical world events can shape the technical landscape. Technology can also shape world events. In an attempt to help Air Force decision makers determine technologies that will best serve their needs in shaping world events, Luker and Myers fleshed out eight different future world threat scenarios.[46,47] They divided the threats in two major parts: state actors and non-state actors. For state actors, the Luker assumed the threat would fight with either physical weapons or information-based weapons, and the fight would occur either on a regular battlefield or on an irregular battlefield (assumes foreign soil for the state actor scenarios). For non-state actors, the Myers considered that the fight might occur either on US or foreign soil, and the adversary might use either physical weapons or information-based weapons.

Different computer vision technologies are applicable in each threat case. For example, if attacks are on foreign soil, computer vision solutions would likely concentrate on improving intelligence, surveillance, and reconnaissance (ISR). For a domestic threat, solutions would more likely concentrate on biometrics. Especially for domestic application of computer vision, human rights would play a part in the decision of which technology to use. From a right to privacy standpoint, US citizens are less accepting of ISR assets monitoring their activities than they are of using biometrics to verify their identity. Intelligent surveillance systems could be used against cargo coming into the US to identify suspect shipments.

In the case of state actors using physical materials on a regular battlefield, computer vision applications might concentrate on systems to apply force to the enemy while maintaining the safety of our forces and systems to warn of impending attacks. Example systems are autonomous combat vehicles and ballistic missile warning systems. On a regular battlefield, computer vision may concentrate on finding tanks under trees; whereas in irregular warfare, efforts may concentrate on finding people dispersed throughout a city.

Computer vision is a more potent deterrent when the adversary uses physical-based material weapons. Computer vision deals with extracting information from physical scenes. In information warfare, the scenes are not physical. In the case of information-based weapons, computer vision technology could indirectly find application, since the pattern recognition algorithms used in computer vision may adapt to finding patterns in information-based weapons.

The coupling of the threat matrix to the specific computer vision technologies is a powerful tool for determining what technologies to invest in. For example, if senior leadership determines that attacks on domestic soil were more likely than attacks on foreign soil, the above discussion shows that biometrics becomes relatively more important than ISR technology. The Air Force should shape the computer vision community to concentrate more heavily on biometrics than on ISR. Furthermore, the survey data will give senior leaders an indication as to when the technologies will mature and an indication of possible hurdles. If the needed technology is already predicted to mature in the near-term, could apply resources elsewhere.

Conclusion

This report provided a view of the future for computer vision technology through the eyes of the academic and industry research community. The Air Force 2025 study motivated the topic of computer vision because many of the conceptual systems identified by the study relied on

[46] Joel J. Luker, "State Actor Threats in 2025" (Air University, 2007).
[47] James W. Myers, "Nonstate Actor Threats in 2025: Blue Horizons Scenarios" (Air University, 2007).

computer vision technology not yet available. This report sought academia and industry opinion because the Air Force continues to rely heavily on them for basic research in this area. The four major recommendations based on the survey data were to focus on artificial intelligence, sensor networks, data fusion, and professionals capable of technology transfer.

One of the largest contributions of this research was the gathering and compilation of the expert responses. The survey data resulted in not only a list of projected significant advances but in an estimation of expected maturity and in an estimation of relative difficulty. The data also provided a list of projected technical and non-technical hurdles that might arise. The analysis showed that nearly all the technologies required in the 2025 study would be available in the 2030 timeframe.

In the context of several possible world threat scenarios, coupling the survey data to a threat matrix provides an important tool to senior Air Force leadership. Computer vision plays a bigger role in a material dominant world than in an information dominant world. Computer vision has a big role to play regardless of whether the threat is a state or non-state actor and independent of whether the conflicts are on domestic or foreign soil. Although the requirements are different for each case, they all include advances in ISR.

Finally, the report recommended that the Air Force use grand challenge problems to help shape computer vision advances. Grand challenges help to focus the research community and might be useful when traditional funding methods fail.

APPENDIX A: Additional technologies

The survey participants proposed advances in several technologies in addition to the ones discussed in section three. They are discussed here.

Biometrics

Biometrics "deals with identification of individuals based on their biological or behavioral characteristics."[48] The Biometric Consortium website contains a wealth of information about the latest research and advancements in biometrics. Biometrics includes both verification and identification. Verification is determining whether a person's stated identity is accurate, while identification involves determining who a candidate is from a database of possible candidates.

Candidate biological and behavioral characteristics involving computer vision include iris scans, fingerprints, cooperative face recognition, vascular recognition, and gait recognition. From a commercial-use standpoint, biometrics promises to help improve current access control to internet sites such as banks and for physical access to cars, homes, or offices, among other applications. From a military standpoint, biometrics is useful for intelligence, surveillance, reconnaissance (ISR) and security. A few niche markets have been using biometrics for some time. However, they still have not gained widespread use, nor the ability in most cases to rival human performance.

Biometrics as a replacement for current access controls was one of the technology advances most likely to occur. With the exception of two panelists who thought it would mature in the mid-term, all of them agreed it was a near-term technology. Here are some specific comments from the participants. "[We] may be able to use lower intensity light for iris scans and rely on algorithms and computational power to make up for the lower quality images." "Methods for performing [human identification and biometrics] recognition will be as accurate as is possible given the specific input image data. e.g., all aspects of a fingerprint image will be used." "Face recognition will be able to handle very large differences in pose, lighting, and facial expression."

One participant also warned that biometric systems will have to continue to improve: "Yet, there will be an intense "cat and mouse" game between those using biometrics and those wanting to break the systems, and there will be continual evolution of systems which are multi-cue and multimodal, raising the cost and effort needed to spoof systems."

With the "cat and mouse game" in mind, the survey moderator asked the participants to give estimates specifically about non-cooperative face recognition. The survey participants were noticeably less optimistic than they were for the general biometrics case. In fact, of the 35 technologies, this was one of the least likely to be mature by 2030, with 4 panelists expecting the technology to mature after 2030. Five participants estimated mid-term maturity. The remaining two respondents determined this to be a long-term capability.

Face Detection

Face detection involves locating human faces in a scene. Face detection has application to automatic management of image databases. For the Air Force, this technology would be useful for the ISR and security communities. When asked when face detection would likely mature, there was not a clear consensus. Two participants thought it was a near-term technology, four thought it was a mid term technology, and two thought it was long-term. Three determined that it would not be mature by 2030 at all. On how the technology would mature, one survey participant said, "Face detection will continue to mature in a manner similar to speech

---

[48] Anil K. Jain et al., *Biometrics: Personal Identification in Networked Society*, The Kluwer International Series in Engineering and Computer Science ; Secs 479 (New York: Kluwer, 2002).

recognition. It will be slow but steady progress, but by 2030 it will be effective for complex (but fixed) scenes with general groups of people."

Automated Mapping

Another area of predicted advancement was in automated mapping. One panelist predicted that computer vision would be used to update maps. The automated system would be able to recognize relevant changes and create descriptions of the changes by relating previous maps and images to new images. The majority of the panelists believed this was a near-term capability. The Air Force could benefit from this technology for updating maps in remote areas where current maps are not readily available or reliable. Additionally, the change detection used for map updates would also be useful in surveillance and reconnaissance work.

Visual Aids for the Blind

Visual aids for the blind are a special case of human-computer interfaces. While helping blind people to regain their vision is not one of the main missions of the Air Force, advances in this field would indirectly help the Air Force by improving computer-to-human interfaces. About 7 people said that in the mid-term, visual aids would be available to help blind people with mobility, reading, and locating and identifying people. Three people thought this was a long-term capability, and one person thought there was a slight chance this would take longer than 2030. Interestingly, this was one of only four of the 35 predictions where no one thought it was a near-term capability. An artificial eye that actually works would be a more difficult capability to achieve. Again, no one thought this was a near-term capability. Five people thought it was a long-term, capability, three thought it was mid-term, and three thought it take longer than 2030.

Scene Reconstruction

Closely related to both object recognition and sensor improvement is scene reconstruction. One participant predicted that by 2030, we would have an easy way of capturing 3D dynamic scenes. All participants agreed, with 6.5 votes for near-term, 2.5 votes for mid-term, and two votes for long-term. When asked more specifically about capturing 3D structures by determining shape from motion, votes were very similar. This time 5.5 voted for near-term, 3.5 voted for mid-term, and two voted for long term.

Another participant predicted the maturity of real-time computer vision analysis and 3D reconstruction from using camera networks. The panel was evenly split: 3.5 votes for near-term, 3.5 votes for mid-term, and four votes for long-term. When a similar question was asked, but without the requirement for real-time analysis, the panel was slightly more optimistic. Five thought this would mature in the near-term, three thought in the mid-term, two voted for long-term, and a single participant estimated anywhere from near to long-term depending on the sophistication.

Medical Diagnostics

Medical image processing is a major sub-field within image processing. While medical imaging is not generally of direct use to the Air Force, many of the algorithms have direct parallels to military applications. For example, the same algorithms that are used in magnetic resonance imaging have parallels to radar image formation and processing. One panelist predicted that advances would allow for automated differentiation or change detection when studying pathological material. Five participants thought this would be a near-term achievement, three thought it was more mid-term, one thought it was long-term, and one thought it would be anywhere from near to long-term. Another participant predicted that computer-aided diagnostics would perform detection functions, such as for mammography. As additional rational he stated,

“I believe that there is an algorithm … that has been approved for clinical use in Europe. Unfortunately, I don't think it has really taken off yet because European law requires that two radiologists read each study, which reduces the economic value.” Five panelists estimated this to be a near-term technology and five thought it was more mid-term. An additional panelist thought it would be anywhere from near to long-term.

APPENDIX B: Study Results in Table Form

Table 1: Combined results of the Delphi study.

The left-most column shows the question. Columns 2 through 12 show the experts' responses. S indicates soon (by 2014), m stands for mid-term (between 2015 and 2022), l indicates long-term (2023-2030), v means very long-term (beyond 2030), and I indicates impossible. The last columns total the indications in each time period for each question, and the last rows of the table show the totals by panel expert.

	Expert											Totals					
	9	10	11	3	8	6	1	7	5	4	2	soon	mid-term	long-term	past 2030	imposs	no-ans
Industrial inspection	s	s	s	s	s	s	s	s	s	s	s	11.0	0.0	0.0	0.0	0.0	0.0
Low-cost human motion capture enabling device-free controllers for games like Wii	s	s	s	na	m	s	s	s	s	s	s	9.0	1.0	0.0	0.0	0.0	1.0
The use of biometry to replace access control (login, banking, physical access to cars-home-office,...). Includes improvements in biometrics: iris scans, fingerprints, co-operative face recognition (confirming that the subject is the one he/she claims to be), etc.	s	s	s	m	s	m	s	s	s	s	s	9.0	2.0	0.0	0.0	0.0	0.0

Automated, adaptive camera calibration including photometric considerations will be solved.	s	s	S	m	s	s	s	na	s	m	m	7.0	3.0	0.0	0.0	0.0	1.0
Computers will be ubiquitous, embedded in almost all everyday objects.	s	s	S	m	l	s	l	s	s	s	s	8.0	1.0	2.0	0.0	0.0	0.0
Extremely large format (gigapixel) "video" sensors	s/m	s	S	m	l	s	s/m	s	s	s	m	7.0	3.0	1.0	0.0	0.0	0.0
Automated mapping. This primarily would be updating, not new maps. Recognition of relevant changes, creation of descriptions of the changes, relating maps to new images.	s	s	M	m	m	s	s	m	s	s	m/l	6.0	4.5	0.5	0.0	0.0	0.0
Miniature imaging sensors with on-board computation enabling novel surveillance apps	s	m	M	m	s	s	m	m	s	s	s/l	5.3	5.3	0.3	0.0	0.0	0.0

Computer aided diagnostics for human clinical applications such as mammography—detection	s	s	S	m	m	m	m	s	s	m	s/l	5.3	5.3	0.3	0.0	0.0	0.0
Easy way of capturing 3D dynamic scenes.	s	s	s/m	m	l	s	l	s	s	m	s	6.5	2.5	2.0	0.0	0.0	0.0
Interactive zoom/pan/tilt over television/Internet for unlimited number of users of both video and audio.	s	s	S	na	l	s	m	m	s	m	m	5.0	4.0	1.0	0.0	0.0	1.0
Computer aided for human clinical applications—differentiation or change detection in pathological material	s	s	Na	s	s	m	m	m	s	l	s/l	5.3	3.3	1.3	0.0	0.0	1.0
Robust implementations of shape from motion enabling 3d structure capture	s	s	s/m	m	s	l	l	m	s	m	s	5.5	3.5	2.0	0.0	0.0	0.0
The advent of decentralized Computer Vision,	s	s	m	l	s	m	l	m	s	s	s/l	5.3	3.3	2.3	0.0	0.0	0.0

where many intelligent cameras communicate to generate better descriptions through complementary processing.																	
Handwriting in various alphabets. Research in this area did not become popular until the 1990s, so there is a fair amount left to be done. Recognizers can also take advantage of increased computational power.	s	na	S	l	m	s	na	s/l	m	s	l	4.3	2.3	2.3	0.0	0.0	2.0
Real time Computer Vision analysis and 3D reconstruction from camera networks will become the reality.	s	s	s/m	m	s	l	l	l	l	m	m	3.5	3.5	4.0	0.0	0.0	0.0
Stereoscopic television/Internet movies.	s/m	s	M	m	s	v	m	m	s	m	m	3.5	6.5	0.0	1.0	0.0	0.0

Systems for locating and managing networked image and video content will be fully developed. So you will be able to find pictures and videos of events and general locations and with specific participants.	s	s	M	s	m	m	l	s	s	v	l	5.0	3.0	2.0	1.0	0.0	0.0
Natural/comfortable glassesless 3D visual display.	m	l	M	na	s	l	s	v	s	m	s	4.0	3.0	2.0	1.0	0.0	1.0
"Natural human-computer interface" using vision and speech.	s	l	M	m	m	l	m	v	s	m	m	2.0	6.0	2.0	1.0	0.0	0.0
Visual aids for the blind -- including mobility, reading, locating and identifying people.	m	m	M	l	m	m	l	m	l	m	m/v	0.0	7.3	3.3	0.3	0.0	0.0
Automatically annotating large image databases at a semantic level, thus allowing for high-speed, reliable	m	m	m	l	l	l	l	m	s	l	l	1.0	4.0	6.0	0.0	0.0	0.0

image retrieval.																	
Progress toward human-like automatic target recognition, especially when information is available from multiple sensors.	s	l	M	l	v	m	l	m	s	m	m	2.0	5.0	3.0	1.0	0.0	0.0
Intelligent surveillance and monitoring.	s	s	s/m	m	s	m	m	s/l	v	v	s/l	4.2	4.2	0.7	2.0	0.0	0.0
Persistent surveillance - If Moore's Law holds up, then by 2030, there will be camera systems which can watch public locations and identify individuals that are present. This can be used for a wide variety of desirable and perhaps undesirable purposes.	m	s	L	l	s	m	l	l	s	l	v	3.0	2.0	5.0	1.0	0.0	0.0
Intelligent Transportation Systems, including	m	s	v	m	m	m	l	m	v	l	m	1.0	6.0	2.0	2.0	0.0	0.0

self-driving vehicles and intelligent highways--along freeways for long stretchs.																	
Cameras with integrated sensing and processing to produce not (only) images, but interpretations. This in turn will enable a number of fundamental changes, including ubiquitous "robots" with sophisticated vision interacting with people in everyday scenarios.	m	na	s/m	l	m	m	m	v	v	l	m	0.5	5.5	2.0	2.0	0.0	1.0
Self-configuring machine vision systems for industrial, research, and navigation tasks.	m	s	S	na	m	v	l	v	na	l	l	2.0	2.0	3.0	2.0	0.0	2.0
Significant progress in closing the "performance gap vs. human vision"	na	l	m	m	m	l	v	s	l	v	l	1.0	3.0	4.0	2.0	0.0	1.0

for category-level object recognition in natural, cluttered scenes for visible (non-occluded) objects, a limited number of categories and simple spatial configurations.																	
Intelligent Transportation Systems, including self-driving vehicles and intelligent highways--on public roads in mixed traffic.	m	s	v	l	l	m	l	l	v	v	l	1.0	2.0	5.0	3.0	0.0	0.0
Artificial eyes for the blind that actually work.	m	l	l	l	l	m	v	l	v	m	v	0.0	3.0	5.0	3.0	0.0	0.0
Face detection will continue to mature is a manner similar to speech recognition. It will be slow but steady progress, but by 2030 it will be effective for	m	m	s	s	v	l	m	m	i	l	v	2.0	4.0	2.0	2.0	1.0	0.0

complex (but fixed) scenes with general groups of people.																	
Non-cooperative face recognition	m	M	m	l	v	l	m	m	v	v	v/i	0.0	5.0	2.0	3.5	0.5	0.0
Visual understanding for humanoid robots.	m/l	Na	l	l	m	v	m	v	v	v	v	0.0	2.5	2.5	5.0	0.0	1.0
Visual thinking systems capable of artistic abstraction, association, and visual creativity.	m/l	V	l	v	v	v	v	v	s	v	v/i	1.0	0.5	1.5	7.5	0.5	0.0
Soon	20.0	21.0	13.5	4.0	12.0	10.0	6.5	9.7	22.0	9.0	8.7						
Mid-term	13.0	5.0	14.5	15.0	12.0	13.0	11.5	13.7	1.0	12.0	11.5						
Long-term	1.0	5.0	4.0	11.0	7.0	8.0	13.0	4.7	3.0	7.0	8.5						
Very long-term	0.0	1.0	2.0	1.0	4.0	4.0	3.0	6.0	7.0	7.0	5.3						
Impossible	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0						
no answer	1.0	3.0	1.0	4.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0						

APPENDIX C: Combined Round 1 Responses for Question 1
(Typographical errors are artifacts of the survey participant responses.)

Computer aided diagnostics for human clinical applications such as momography. I believe that there is an algorithm call R2 that has been approved for clinical use in Europe. Unfortunately, I don't think it has really taken off yet because European law requires that two radiologist read each study, which reduces the economic value.
However, it is just a matter of time before this becomes important in the US.

Persistant survalence - If Moore's Law holds up, then by 2030, there will be camera systems which can watch public locations and identify individuals that are present. This can be used for a wide variety of desirable and perhaps undesirable purposes. Face detection will continue to mature is a manner similar to speech recognition. It will be slow but steady progress, but by 2030 it will be effective for complex (but fixed) scenes with general groups of people.

Systems for locating and managing personal and social networked image and video content will be fully developed. So you will be able to find pictures and videos of events and general locations and with specific participants. This will become a more common form of interpersonal communication.

Artificial eyes for the blind that actually work.

Self-configuring machine vision systems for industrial,research, and navigation tasks.

Visual thinking systems capable of artistic abstraction, association, and visual creativity.

Miniature imaging sensors with on-board computation enabling novel surveillance apps

Robust implementations of shape from motion enabling 3d structure capture

Low-cost human motion capture enabling device-free controllers for games like Wii

Extremely large format (gigapixel) "video" sensors

Content based retrieval of imagery on the web

Handwriting in various alphabets.
WHY: Research in this area did not become popular until the 1990s, so there is a fair amount left to be done. Recognizers can also take advantage of increased computational power.

Improvements in biometrics: iris scans, fingerprints, co-operative face recognition (confirming that the subject is the one he/she claims to be), etc.
WHY: For example, we may be able to use lower intensity light for iris scans and rely on algorithms and computational power to make up for the lower quality images.

Real time Computer Vision analysis and 3D reconstruction from camera networks will become the reality.

The sub problems of Automated, adaptive camera calibration including photometric considerations will be solved.

Reconstruction: Integrated approaches for reconstructing 3-D geometry of large scale scenes will be available. The theoretical, algorithmic, and implementation issues associated with inferring 3-D shape from one or more images or image sequences will be well understood and integrated, and they will exploit all available cues (correspondence, shading, shadows, lighting).
On the one hand, they will be robust to all imaging situations in the sense of returning the meaningful solutions under all conditions, while modeling limitaitons of the reconstruciton (e.g., accuracy and ambiguities) will be explicit. They will be able to handle a broad range of materials (from matte to glossy, from translucent to opaque), complex geometries and broad range of scales.
For underconstrained problems, priors about specific objects or object classes will be incorporated into reconstruction methods in an accesible fashion.

Specific object recognition will be very robust. It will be possible to "train" a system by showing one or more examples of the objects in an unsegmented scene and the object will be recognizable in new images taken under a broad range of conditions (pose, lighting, differing shape configurations, etc.).

Recognition of previously unseen objects: Today, this is referred to as generic object recogntion or recognition of object classes. By 2030, it will better understood how to describe or identify a previously unseen object in terms of a great deal of prior knowledge about a very large number of broad object classes. A large scale ontology of objects and scenes will have been developed, and given one or more images, recognition methods will describe the image content in terms of this ontology.

Human identification and biometrics--Methods for performing recognition will be as accurate as is possible given the specific input image data. e.g., all aspects of a fingerprint image will be used, not just say minutiae and face recognition will be able to handle very large differences in pose, lighting, and facial expression. This limits on recognition performance will not be algorithimic, but rather the fundamental accuracy of the specific biometric trait given the within class variation of that trait for an individual and the between class variation for the specific population.
Yet, there will be an intense "cat and mouse" game between those using biometrics and those wanting to break the systems, and there will be continual evolution of systems which are multi-cue and multimodal, raising the cost and effort needed to spoof systems.

Autonomous vehicle and robot

Biometric identification and verification

Visual surveillance

Industrial inspection, medical image analysis

Automatic driving of passenger cars over long stretches, especially along freeways.

Stereoscopic television/Internet movies.

Interactive zoom/pan/tilt over television/Internet for unlimited number of users of both video and audio.

Visual aids for the blind -- including mobility, reading, locating and identifying people.

Automated mapping. This primarily would be updating, not new maps. Recognition of relevant changes, creation of descriptions of the changes, relating maps to new images.

Intelligent surveillance and monitoring.

Automated driving on public roads in mixed traffic.

"Natural human-computer interface" using vision and speech. (By 2030, computers will be ubiquitous, embedded in almost all everyday objects.)

3D Television/Video: (a) Easy way of capturing 3D dynamic scenes. (b) Natural/comfortable glassesless 3D visual display.

Visual understanding for humanoid robots.

Artificial human eye.

Intelligent Transportation Systems (including intelligent perhaps self-driving vehicles and intelligent highways), where vision plays a key role.

Ease in searching for multimedia (esp. images and video) data from huge databases.

The advent of "Intelligent Cameras", i.e. cameras with integrated sensing and processing to produce not (only)images, but interpretations.
This in turn will enable a number of fundamental changes, including ubiquitous "robots" with sophisticated vision interacting with people in everyday scenarios.

The advent of decentralized Computer Vision, where many intelligent cameras communicate to generate better descriptions through complementary processing.
This will enable applications such as the truly "Intelligent" Home, and persistent tracking and surveillance.

The use of biometry to replace access control (login, banking, physical access to cars-home-office,...)

Face recognition

Autonomous vehicles (e.g., cars that drive themselves) image-based indexing/search of image databases (and the web)

Significant progress in closing the "ROC gap" with human vision for category-level object recognition in natural, cluttered scenes for visible (non-occluded) objects, a limited number of categories and simple spatial configurations.

Corresponding progress in automatically annotating large image databases at a semantic level, thus allowing for high-speed, reliable image retrieval.

Corresponding progress in detecting and differentiating tumors and other pathological structures in medical images.

Corresponding progress in automatic target recognition, especially when information is available from multiple sensors.

APPENDIX D: Combined Round 1 Responses for Question 2
(Typographical errors are artifacts of the survey participant responses.)

Dimensionality reduction - Right now there is no systematic method for finding sparse representations of data which provide the key information necessary to make decisions and estimate unknown quantities.

Computationally limited inference - There is a lack of a systematic framework or theory for making decisions and inferring relationships when the available data is enormous and the computational resources are limited. This problem will need to be solved or at least addressed.

Automated discovery of relationships - Right now pattern recognition methods work well if one knows what he or she is looking for, but there is no systematic approach for unstructured discovery from data. Humans are very good at discovering patterns from information they can understand, but they can not effectively process very high dimensional data with complex structure such as high dimensional graphs. Data mining is an attempt to address this problem, but so far the work that has done falls far short.

Device Technology improvements

Comprehensive theory of machine learning that works.

Computational theory of creativity.

Computer vision apps must be made robust to real-world conditions

Further improvement in SfM techniques

Vision algorithms must be designed to take advantage of the persistence of their sensors and learn from their experiences

Make better use of context

Harnessing the power of distributed computing. Google does that now but we will need new software models for pictorial pattern recognition.
We should be able to follow different interpretations of an image in parallel and have the processors communicate with each other so that at the end we obtain a form of a consensus. Even while the computation goes on, processors may be moved from one interpretation to another that seems more promising.

Automated adaptive camera calibration including photometric changes.

Reconstruction -- fully developed set of physical and mathematicla models for reconstruction. Robust and/or optimal algorithms for reconstruction, including prior information about objects, scenes being reconstructed.

Segmentation -- It will be critical to be able to segment objects to be recognized, but its not clear if segmentation will be distiinct from recognition. Will bottom up segmentation be a distinct process, or will it be integrated with and a consequence of recognition.

Acquisition and organization of object/scene/image information into an ontolgoy. There has been scant real need for this at the present stage of recognition, but as recognition becomes more capable it will be more critical. There is early work stemming from two decades of classical AI that may become more relevant.

Compact, accessible, effective models of object appearance -- we will need advances in the way to represent and learn models of objects which can support a broad range of vision applications.

There was no fundamental breakthrough or leap-forward in computer vision in the past decade, and I foresee that there will not be one in the next two decades either. Advancements are cumulative and require years of hard work and effectively incorporating information from modalities other than visual light camera.

Better Stereo and Depth from Motion.

Integration of multiple-sensor modalities.

Redefinition of zoom, and then progress toward it -- what is currently called zoom should be called image scaling.

Higher quality omni-directional video.

Consistent and stable image analysis with varying lighting conditions (weather, shadows, sun angle).

Learning techniques to improve techniques -- that both find more general solutions (or descriptions) and specialize general descriptions to important variations.

Better representations of real-time events.

Reliable representations of object classes.

Standards need to be defined to allow seamless integration across platforms, software and middleware modules.

Both algorithmic and processor advances are needed to achieve this vision

Associative memory

Massively parallel computers/computation

Full understanding of human visual system

Overcome the "context vs. computation" dilemma in computer vision.
Models which accommodate context (e.g., context-sensitive grammars) are computationally intractable whereas computationally efficient methods (e.g., those based on coarse-to-fine representation and search) do not accommodate context, and hence are ultimately limited in selectivity.

Learning contextual models from reasonably-sized training sets. As the complexity of the space of allowed interpretations increases, beginning to match that of human descriptions, the number of samples per interpretation will inevitably be small, requiring highly efficient learning algorithms, particularly if disambiguation is based on contextual constraints.

APPENDIX E: Consolidated Round 1 Responses to Question 3
(Typographical errors are artifacts of the survey participant responses.)

If Moore's Law breaks down, it could change everything. If silicon hits a dead end, then I don't believe that nanotechnology is going to save the day. This isn't nontechnical, but it is outside the realm of what researchers in imaging and pattern recognition can control.

Society may decide it doesn't want technologies that seem to socially invasive. This could lead to some government regulation of technologies similar to what we have seen in encryption and stem-cell research.

Abandonment of respect for intellectual achievement in favor of religious fundamentalism.

War

Failure to protect the environment, and in turn, loss of our prosperous lifestyle and ability to innovate efficiently.

Excessive privacy restrictions limiting both research and fielding of surveillance products

The biggest obstacle in the past has been overly optimistic promises that lead to disappointment by the research sponsors.

Automatic decision systems will make mistakes, and while they will be
more accurate than humans, I don't think that society will be accepting of the frequencies of these mistakes. While " to err is human,"...
Consider accidents caused by autonomous vehicles or today's driver's
aids (parking, cruise control, driver monitoring systems).

Large-scale systems will be shown to have some sort of statistical bias, which were not deliberately introduced, but which might be seen as benefiting one demographic vs. others. E.g., Consider if biometrics were more effective for one race, gender, ethnicity than for another.

Very often, there does not exist strong support from industry, which is crucial for computer vision in order to further succeed. In other words, industry so far does not think computer vision is worth investment.
Therefore, most computer vision projects are conducted in universities, and financed by NSF or military.

Economic: Willingness to support the necessary research.

Cultural acceptance of automated driving.

It would take a major upheaval to stop this train, but it's possible…A sudden global recession due to global warming, possibly...

Slowdown of progress in computers.

Lack of funding in Computer Vision research.

BIBLIOGRAPHY

"Air Force 2025." Air University, 1996.

Ciufo, Chris A. "COTS: 10 Years after - Well, Sure…but What About the Next 10 Years?" *Military Embedded Systems* (2006).

Comyn, Paddy. "Sensing Forward to a Driverless Future." *The Irish Times* 21 February 2007.

Dakley, Norman C. *The Delphi Method : An Experimental Study of Group Opinion*. Santa Monica, Calif.: Rand Corporation, 1969.

DARPA. "DARPA Grand Challenge 2005." http://www.grandchallenge.org/.

Dekleva, Sasa. "Delphi Study of Software Maintenance Problems." Paper presented at the Conference on Software Maintenance, Los Alamitos, CA, USA 1992.

Glenn, Jerome C. and Theodore J. Gordon, ed. *Futures Research Methodology Version 2.0* American Council for the United Nations University Millennium Project, 2003.

Jain, Anil K., Ruud Bolle, Sharath Pankanti, and NetLibrary Inc. *Biometrics: Personal Identification in Networked Society*, The Kluwer International Series in Engineering and Computer Science ; Secs 479. New York: Kluwer, 2002.

James, Geoffrey. "Out of Their Minds." *Red Herring*, 22 August 2002.

Jolson, Marvin A., and Gerald L. Rossow. "The Delphi Process in Marketing Decision Making." *Journal of Marketing Research* VIII (1971): 443-48.

Kurzweil, Ray. *The Singularity Is Near : When Humans Transcend Biology*. New York: Viking, 2005.

Linstone, Harold A., Robert U. Ayres, Ted Gordon, Theodore Modis, Tessaleno C. Devezas, and Joseph F. Coates. "Book Review and Discussion: The Singularity Is Near: When Humans Transcent Biology." *Technological Forecasting and Social Change* 73 (2006): 95-127.

Luker, Joel J. "State Actor Threats in 2025." Air University, 2007.

Mackey, Morgan. "Nanotechnology Applications for ISR: The Solution to the Intelligence Gap? (DRAFT)." Air University, 2007.

Myers, James W. "Nonstate Actor Threats in 2025: Blue Horizons Scenarios." Air University, 2007.

"NHAA Journal." http://www.cs.cmu.edu/afs/cs/user/tjochem/www/nhaa/Journal.html.

NIST. "NIST Face Recognition Grand Challenge." http://face.nist.gov/frgc.

"No Hands across America Home Page." http://www.cs.cmu.edu/afs/cs/user/tjochem/www/nhaa/nhaa_home_page.html.

Okoli, Chitu, and Suzanne D. Pawlowski. "The Delphi Method as a Research Tool: An Example, Design Considerations and Applications." *Information and Management* 42 (2004): 15-29.

Park, Insu, Jinkyu Lee, H. Raghav Rao, and Shambhu J. Upadhyaya. "Guest Editorial:Part 2: Emerging Issues for Secure Knowledge Management—Results of a Delphi Study." *Systems, Man and Cybernetics, Part A, IEEE Transactions on* 36, no. 3 (2006): 421-28.

Phillips, P. Jonathon, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, and William Worek. "Recognition Grand Challenge Results." Paper presented at the 7th International Conference on Automatic Face and Gesture Recognition 2006.

Rowe, Gene, and George Wright. "The Delphi Technique as a Forecasting Tool: Issues and Analysis." *International Journal of Forecasting* 15 (1999): 353-75.

Saunders, Gregory. "COTS in Military Systems: A Ten Year Perspective." In *Military & Aerospace Electronics Show*. Baltimore, MD, 2004.

Simons, Anna, Scott Chapman, James Engle, Meyer Kestnbaum, Andrew Krepinevich, R. W. Lanham, and Daniel McClellan. "The Military Officer in 2030: Secretary of Defense 2003 Summer Study." Director of Net Assessment, Office of the Secretary of Defense, 2003.

Stockman, George, and Linda G. Shapiro. *Computer Vision*: Prentice Hall PTR, 2001.